

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-293764

(43)Date of publication of application : 04.11.1998

(51)Int.Cl.

G06F 17/30

(21)Application number : 09-102159

(71)Applicant : FUJITSU LTD

(22)Date of filing : 18.04.1997

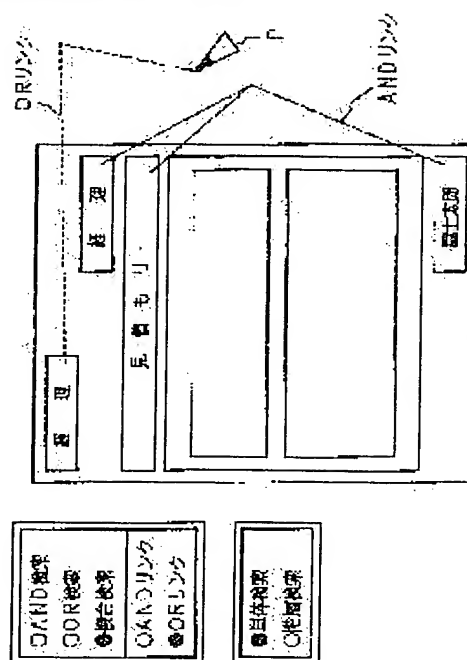
(72)Inventor : GOTO MASATOMO

## (54) STRUCTURED DOCUMENT DATABASE RETRIEVING METHOD, STRUCTURED DOCUMENT DATABASE RETRIEVAL SYSTEM AND RECORDING MEDIUM

## (57)Abstract:

**PROBLEM TO BE SOLVED:** To provide a retrieving method of a structured document database which easily designates an element of an object to be retrieved, easily inputs a keyword to the element and performs retrieval in a short time even at the time of retrieving a structured document that is unknown to a user in a retrieving method to a database which accumulates element names that constitute document and structured documents in which the hierarchical structure of an element is defined.

**SOLUTION:** This device shows a document format diagram that defines layout positions of contents of an element, accepts an input of a keyword that is associated with the layout position of the document format diagram, identifies an element that is an object to be retrieved through the layout position where a keyword is inputted, retrieves the content of the identified element based on an inputted keyword and acquires a document.



## LEGAL STATUS

[Date of request for examination] 13.03.2001

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平10-293764

(43) 公開日 平成10年(1998)11月4日

(51) Int.Cl.<sup>6</sup>

G 0 6 F 17/30

識別記号

F I

G 0 6 F 15/403

15/40

15/403

3 2 0 A

3 7 0 A

3 2 0 Z

審査請求 未請求 請求項の数13 O L (全 12 頁)

(21) 出願番号 特願平9-102159

(22) 出願日 平成9年(1997)4月18日

(71) 出願人 000005223

富士通株式会社

神奈川県川崎市中原区上小田中4丁目1番  
1号

(72) 発明者 後藤 正智

神奈川県川崎市中原区上小田中4丁目1番  
1号 富士通株式会社内

(74) 代理人 弁理士 河野 登夫

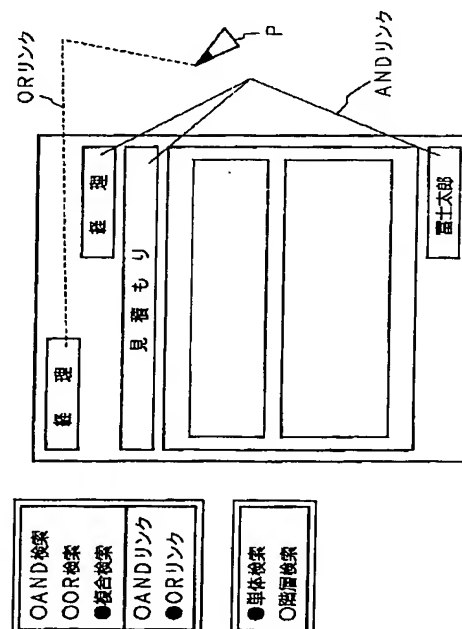
(54) 【発明の名称】 構造化文書データベース検索方法、構造化文書データベース検索システム及び記録媒体

(57) 【要約】

【課題】 SGML (Standard Generalized Markup Language : ISO 8879) で記述された文書のように、文書を構成するエレメントの名称、及びエレメントの階層構造が定義された構造化文書が蓄積されているデータベースに対する検索方法に関し、文書構造がユーザにとって未知である構造化文書を検索する場合であっても、検索対象のエレメントの指定、及びそのエレメントに対するキーワードの入力が容易であって、検索が短時間でできる構造化文書データベースの検索方法を提供する。

【解決手段】 エレメントの内容のレイアウト位置を規定した文書フォーマット図を表示し、文書フォーマット図のレイアウト位置に対応付けたキーワードの入力を受け付け、キーワードが入力されたレイアウト位置によって検索対象のエレメントを識別し、入力されたキーワードに基づいて、識別したエレメントの内容を検索し、文書を獲得する。

スタイル情報を用いたキーワード入力例の図



## 【特許請求の範囲】

【請求項 1】 文書を構成するエレメントの名称、及びエレメントの階層構造が定義され、文書におけるエレメントの内容の表示方法を規定する文書の成形規則が関連付けられた構造化文書が構造的に分割されて蓄積されている構造化文書データベースに対する検索方法において、

エレメントの内容のレイアウト位置を規定した文書フォーマット図を表示し、

文書フォーマット図の前記レイアウト位置に対応付けた 10 キーワードの入力を受け付け、

キーワードが入力されたレイアウト位置によって検索対象のエレメントを識別し、

入力されたキーワードに基づいて、識別したエレメントの内容を検索し、文書を獲得することを特徴とする構造化文書データベース検索方法。

【請求項 2】 文書を構成するエレメントの名称、及びエレメントの階層構造が定義された構造化文書が構造的に分割されて蓄積されている構造化文書データベースに対する検索方法において、

所定の階層に位置するエレメント名を、その下位の階層情報の表示情報とともに表示し、

表示したエレメント名より下位の階層に位置する、エレメント名を表示すべき階層の指定を受け付け、

指定された階層のエレメント名を、その階層関係を表す情報とともに表示し、

表示したエレメント名に対応付けたキーワードの入力を受け付け、

キーワードが入力されたエレメント名によって検索対象のエレメントを識別し、

入力されたキーワードに基づき、識別したエレメントの内容を検索し、文書を獲得することを特徴とする構造化文書データベース検索方法。

【請求項 3】 文書を構成するエレメントの名称、及びエレメントの階層構造が定義された構造化文書が構造的に分割されて蓄積されている構造化文書データベースに対する検索方法において、

検索対象のエレメントを含む文書を構成する全てのエレメントのエレメント名を用いて、エレメントの階層構造を示す構造図を表示し、

該構造図のエレメント名に対応付けたキーワードの入力を受け付け、

キーワードが入力されたエレメント名によって検索対象のエレメントを識別し、

入力されたキーワードに基づいて、識別したエレメントの内容を検索し、文書を獲得することを特徴とする構造化文書データベース検索方法。

【請求項 4】 文書を構成するエレメントの名称、及びエレメントの階層構造が定義され、文書におけるエレメントの内容の表示方法を規定する文書の成形規則が関連 50

2

付けられた構造化文書が構造的に分割されて蓄積されている構造化文書データベースに対する検索方法において、

キーワードの入力方法の指定を受け付け、

指定された入力方法に応じて、

エレメントの内容のレイアウト位置を規定した文書フォーマット図を表示し、

文書フォーマット図の前記レイアウト位置に対応付けたキーワードの入力を受け付け、

キーワードが入力されたレイアウト位置によって検索対象のエレメントを識別し、

又は、

所定の階層に属するエレメント名を、その階層関係を表す情報とともに表示し、

表示したエレメント名及びその階層関係を表す情報に基づく、エレメント名を表示すべき階層の指定を受け付け、

指定された階層のエレメント名を、その階層関係を表す情報とともに表示し、

20 表示したエレメント名に対応付けたキーワードの入力を受け付け、

キーワードが入力されたエレメント名によって検索対象のエレメントを識別し、

又は、

検索対象のエレメントを含む文書を構成する全てのエレメントのエレメント名を用いて、エレメントの階層構造を示す構造図を表示し、

該構造図のエレメント名に対応付けたキーワードの入力を受け付け、

30 キーワードが入力されたエレメント名によって検索対象のエレメントを識別し、

指定された入力方法によって識別したエレメントの内容を、入力されたキーワードに基づいて検索し、文書を獲得することを特徴とする構造化文書データベース検索方法。

【請求項 5】 識別したエレメントに加え、該エレメントの下位の階層に属するエレメントを検索対象とする請求項 1 乃至 4 のいずれかに記載の構造化文書データベース検索方法。

40 【請求項 6】 検索対象のエレメントが複数の場合、AND 検索又は OR 検索のいずれかの検索条件で検索を実行する請求項 1 乃至 5 のいずれかに記載の構造化文書データベース検索方法。

【請求項 7】 検索対象のエレメントが複数の場合、AND 検索及び／又は OR 検索の指定を受け付け、指定された検索条件で検索を実行する請求項 1 乃至 5 のいずれかに記載の構造化文書データベース検索方法。

【請求項 8】 文書を構成するエレメントの名称、及びエレメントの階層構造が定義され、文書におけるエレメントの内容の表示方法を規定する文書の成形規則が関連

3

付けられた構造化文書が構造的に分割されて蓄積されている構造化文書データベースの検索システムにおいて、エレメントの内容のレイアウト位置を規定した文書フォーマット図を表示する手段と、  
文書フォーマット図の前記レイアウト位置に対応付けたキーワードの入力を受け付ける手段と、  
キーワードが入力されたレイアウト位置によって検索対象のエレメントを識別する手段と、  
入力されたキーワードに基づいて、識別したエレメントの内容を検索し、文書を獲得する手段とを備えたことを特徴とする構造化文書データベース検索システム。

【請求項 9】 文書を構成するエレメントの名称、及びエレメントの階層構造が定義された構造化文書が構造的に分割されて蓄積されている構造化文書データベースの検索システムにおいて、  
所定の階層に位置するエレメント名を、その下位の階層情報の表示情報とともに表示する手段と、  
表示したエレメント名より下位の階層に位置する、エレメント名を表示すべき階層の指定を受け付ける手段と、  
指定された階層のエレメント名を、その階層関係を表す情報とともに表示する手段と、  
表示したエレメント名に対応付けたキーワードの入力を受け付ける手段と、  
キーワードが入力されたエレメント名によって検索対象のエレメントを識別する手段と、  
入力されたキーワードに基づき、識別したエレメントの内容を検索し、文書を獲得する手段とを備えたことを特徴とする構造化文書データベース検索システム。

【請求項 10】 文書を構成するエレメントの名称、及びエレメントの階層構造が定義された構造化文書が構造的に分割されて蓄積されている構造化文書データベースの検索システムにおいて、  
検索対象のエレメントを含む文書を構成する全てのエレメントのエレメント名を用いて、エレメントの階層構造を示す構造図を表示する手段と、  
該構造図のエレメント名に対応付けたキーワードの入力を受け付ける手段と、  
キーワードが入力されたエレメント名によって検索対象のエレメントを識別する手段と、  
入力されたキーワードに基づいて、識別したエレメントの内容を検索し、文書を獲得する手段とを備えたことを特徴とする構造化文書データベース検索システム。

【請求項 11】 文書を構成するエレメントの名称、及びエレメントの階層構造が定義され、文書におけるエレメントの内容の表示方法を規定する文書の成形規則が関連付けられた構造化文書が構造的に分割されて蓄積されている構造化文書データベースに対して検索を実行するコンピュータでの読み取りが可能なプログラムコード手段を格納しており、前記コンピュータでの使用が可能な記録媒体において、

4

前記プログラムコード手段が、  
前記コンピュータに、エレメントの内容のレイアウト位置を規定した文書フォーマット図を表示させるコード手段と、  
前記コンピュータに、文書フォーマット図の前記レイアウト位置に対応付けたキーワードの入力を受け付けさせるコード手段と、  
前記コンピュータに、キーワードが入力されたレイアウト位置によって検索対象のエレメントを識別させるコード手段と、  
前記コンピュータに、入力されたキーワードに基づいて、識別したエレメントの内容を検索させ、文書を獲得させるコード手段とを含むことを特徴とする記録媒体。

【請求項 12】 文書を構成するエレメントの名称、及びエレメントの階層構造が定義された構造化文書が構造的に分割されて蓄積されている構造化文書データベースに対して検索を実行するコンピュータでの読み取りが可能なプログラムコード手段を格納しており、前記コンピュータでの使用が可能な記録媒体において、

前記プログラムコード手段が、  
前記コンピュータに、所定の階層に位置するエレメント名を、その下位の階層情報の表示情報とともに表示させるコード手段と、  
前記コンピュータに、表示したエレメント名より下位の階層に位置する、エレメント名を表示すべき階層の指定を受け付けさせるコード手段と、  
前記コンピュータに、指定された階層のエレメント名を、その階層関係を表す情報とともに表示させるコード手段と、  
前記コンピュータに、表示したエレメント名に対応付けたキーワードの入力を受け付けさせるコード手段と、  
前記コンピュータに、キーワードが入力されたエレメント名によって検索対象のエレメントを識別させるコード手段と、  
前記コンピュータに、入力されたキーワードに基づき、識別したエレメントの内容を検索させ、文書を獲得させるコード手段とを含むことを特徴とする記録媒体。

【請求項 13】 文書を構成するエレメントの名称、及びエレメントの階層構造が定義された構造化文書が構造的に分割されて蓄積されている構造化文書データベースに対して検索を実行するコンピュータでの読み取りが可能なプログラムコード手段を格納しており、前記コンピュータでの使用が可能な記録媒体において、  
前記プログラムコード手段が、  
前記コンピュータに、検索対象のエレメントを含む文書を構成する全てのエレメントのエレメント名を用いて、エレメントの階層構造を示す構造図を表示させるコード手段と、  
前記コンピュータに、該構造図のエレメント名に対応付けたキーワードの入力を受け付けさせるコード手段と、

10

20

30

40

50

前記コンピュータに、キーワードが入力されたエレメント名によって検索対象のエレメントを識別させるコード手段と、

前記コンピュータに、入力されたキーワードに基づいて、識別したエレメントの内容を検索させ、文書を獲得させるコード手段とを含むことを特徴とする記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、例えば SGML (Standard Generalized Markup Language : ISO 8879) で記述された文書のように、文書を構成するエレメントの名称、及びエレメントの階層構造が定義された構造化文書が蓄積されているデータベースに対する検索方法、検索システム及びそのデータベースに対する検索のコンピュータプログラムが記録されている記録媒体に関する。

【0002】

【従来の技術】アメリカでは、国防総省、多くの学術研究機関がSGMLを導入しており、民生利用も拡大している。また、インターネットで人気が高いWWW (World-Wide Web) で流通しているコンテンツも、HTML (Hyper Text Markup Language) で記述された構造化文書である。また、インターネット/イントラネットの普及により、インターネットの場合は世界中の文書に、イントラネットの場合は会社のようなグループ内の文書にネットワークを通じてアクセスできるようになっている。

【0003】特にイントラネットの普及で、社内文書のような定型文書が電子化されて蓄積されるようになり、ペーパーレス化が促進されている。アメリカ国防総省が、企業活動に関わるすべての情報を電子化してオンライン上で情報交換できるようにしようと展開しているCA LIS (Continuous Acquisition and Life-cycle Support) プロジェクトでは、情報交換にSGMLを採用しているので、構造化文書はより一般的になると考えられる。

【0004】構造化文書は論理構造を持っているため、この論理構造を利用することで検索が高速にしかも容易に行える。即ち、SGML文書は文書型定義 (DTD) により文書の構造が定義され、その構造は図11に示すようなツリー構造で表される。従って、例えば、「レポート」というエレメントが「宛先」「タイトル」「本文」「担当」及び「発信元」のエレメントから構成され、これらの中の「本文」が「概要」及び「詳細」のエレメントから構成されている文書構造を持つ文書群から、「〇〇さん」を宛先として発信されたレポートを検索する場合、全文を検索の対象としなくても、「宛先」のエレメントを指定し、「〇〇さん」をキーワードとして検索すればよい。

【0005】

【発明が解決しようとする課題】しかし、構造化文書において上述のような検索を行おうとする場合、ユーザは、キーワードを与えるときに検索対象とすべき文書構

造を指定しなければならないので構造化文書の文書構造を把握しておく必要がある。

【0006】例えば、ユーザは、図11のような8つのエレメントからなるツリー構造のノード毎に付けられているタグ名 (エレメント名) を把握して、「〇〇さん」に発信されたレポートを検索する場合は、「宛先」というエレメント名が検索対象であることを指定し、「〇〇さん」をキーワードとして入力してやらなければならない。

【0007】従って、このような検索手法は、高速に検索ができるにもかかわらず、構造化文書の文書構造を把握しておかなければならないという負担が大きい。ユーザは、従来の全文検索の手法、リレーショナルデータベースの検索手法等を用いる方が検索が容易であると感ずる。

【0008】本発明はこのような問題点を解決するためになされたものであって、構造化文書の文書構造の図式をエレメント名を用いてユーザに示し、図式上のエレメント名に対してユーザがキーワードを入力したエレメントを対象として検索することにより、文書構造がユーザにとって未知である構造化文書を検索する場合であっても、検索対象のエレメントの指定、及びそのエレメントに対するキーワードの入力が容易であって、検索が短時間でできる構造化文書データベースの検索方法、検索システム及び構造化文書データベース検索のコンピュータプログラムが記録されている記録媒体の提供を目的とする。

【0009】

【課題を解決するための手段】図1は本発明の構造化文書データベース検索方法を実施するシステムの構成図である。表示装置1は、エレメント名を用いた、エレメントの内容のレイアウト位置を規定した文書フォーマット図、構造化文書の文書構造の所定の階層の構造、又はエレメントの階層構造の構造図といった、構造化文書の文書構造の図式と図式上のエレメント名に対するキーワードの入力画面、検索対象のエレメントがキーワードを入力されたエレメントだけであるか、下位の階層のエレメントを含むかを指定し、またOR、AND の検索条件を指定する画面、検索結果をディスプレイに表示する。

【0010】キーボード、マウス等からなる入力装置2は、画面に表示された文書構造の図式上のエレメント名に対してキーワードを入力する手段、またキーワードの入力方法を指定する手段、検索条件を指定する手段、各種入力画面の表示命令を入力する手段、及び検索の実行命令を入力する手段である。

【0011】インタフェースモジュール3は、入力装置2を用いた、ユーザによるキーワード、検索条件等の入力フォームを作成して表示装置1に表示させたり、表示に対する入力装置2からのユーザの入力を受け付け、その入力から問い合わせ (クエリ) 言語を作成してサーチ

エンジン4に渡す。

【0012】サーチエンジン4は渡された問い合わせ言語により、構造化文書が蓄積されているドキュメントDB(データベース)7に対して検索を行う。SGMLパーサ5はDTDを基に、SGML文書を構文解析(パース)してその妥当性を検証し、SGML文書のエンティティツリーを作成する。

【0013】エンティティマネージャ6は構文解析した結果をドキュメントDB7へ格納し、文書の実体を管理する。ドキュメントDB7には、ドキュメントを構成するエレメントの名称及びエレメントの階層構造がDTDにより定義された構造化文書がエレメントに分割されて蓄積されている。上述のSGMLパーサ5、エンティティマネージャ6、及びドキュメントDB7によって構成されるSGMLシステムは、構造化文書を構文解析し、エレメントに分割して格納する機能を提供する。

【0014】スタイルシートDB9は、SGML文書のDTDに応じた文書成形の規則を定めたスタイルシートが蓄積されている。DSSSLエンジン8は、構文解析されたSGML文書を、スタイルシートDB9に蓄積されている、SGML文書のDTDに応じた文書成形の規則を定めたスタイルシートに沿って表示するための文書成形を行う。上述のDSSSLエンジン8及びスタイルシートDB9によって構成されるDSSSLシステムは、SGML文書を表示するための機能を提供する。

【0015】インタフェースモジュール3は、SGMLシステム(5,6,7)から文書構造を読み出し、表示形式に沿って表示装置1に入力フォームを出力し、またDSSSLシステム(8,9)と連携してキーワードを取得する。サーチエンジン4はSGMLシステム(5,6,7)及びDSSSLシステム(8,9)と連携して、与えられたクエリに応じて検索を行う。

【0016】本発明の構造化文書データベース検索方法、構造化文書検索システム及び記録媒体は、インタフェースモジュール3が、ドキュメントDB7に蓄積されているDTD、及びスタイルシートDB9に蓄積されているそのDTDのスタイルシートによって、エレメントの内容のレイアウト位置を規定した文書フォーマット図を、表示装置1に、例えば枠組みで表示し、文書フォーマット図のレイアウト位置の上下左右等、レイアウト位置に対応付けたキーワードの入力を受け付け、キーワードが入力されたレイアウト位置によって検索対象のエレメントを識別する。サーチエンジン4は、入力されたキーワードに基づいて、ドキュメントDB7に蓄積されているドキュメントの中の、インタフェースモジュール3が識別したエレメントの内容を検索し、文書を獲得する。

【0017】また、本発明の構造化文書データベース検索方法、構造化文書検索システム及び記録媒体は、インタフェースモジュール3が、ドキュメントDB7に蓄積されているDTDによって、まずエレメントの階層構造のう

ち、所定の階層に属するエレメント名とその階層関係を表す情報とを表示装置1に表示し、表示したエレメント名及びその階層関係を表す情報とに基づき、エレメント名を表示すべき階層の指定を受け付け、指定された階層のエレメント名を、その階層関係を表す情報とともに表示装置1に表示し、エレメント名に対応付けた、例えばダイアログボックスへのキーワードの入力を受け付け、キーワードが入力されたエレメント名によって検索対象のエレメントを識別する。サーチエンジン4は、入力されたキーワードに基づいて、ドキュメントDB7に蓄積されているドキュメントの中の、インタフェースモジュール3が識別したエレメントの内容を検索し、文書を獲得する。

【0018】さらに、本発明の構造化文書データベース検索方法、構造化文書検索システム及び記録媒体は、インタフェースモジュール3が、ドキュメントDB7に蓄積されているDTDによって、検索対象のエレメントを含む文書を構成する全てのエレメントのエレメント名を用いて、エレメントの階層構造を示す構造図を表示し、この構造図のエレメント名に対応付けた、例えばダイアログボックスへのキーワードの入力を受け付け、キーワードが入力されたエレメント名によって検索対象のエレメントを識別する。サーチエンジン4は、入力されたキーワードに基づいて、ドキュメントDB7に蓄積されているドキュメントの中の、インタフェースモジュール3が識別したエレメントの内容を検索し、文書を獲得する。

【0019】従って、本発明の構造化文書データベース検索方法、構造化文書検索システム及び記録媒体は、文書構造がユーザにとって未知である構造化文書を検索する場合であっても、検索対象のエレメントの指定、及びそのエレメントに対するキーワードの入力が容易であって、検索を短時間で行うことができる。

#### 【0020】

【発明の実施の形態】以下に、本発明の構造化文書データベース検索方法(以下、本発明方法という)を図に基づいて説明する。まず、本発明方法に、前述の文書フォーマット図であるスタイル情報を用いる場合について、図2のキーワード入力の説明図、図3のフローチャート、及び図4のキーワード入力例の図に基づき説明する。

【0021】SGMLでは文書型定義(DTD)で文書の型、即ち、章、節、項といった文書構造を定義する。インタフェースモジュール3は、ドキュメントDB7に格納されているSGML文書のDTDを読み出し、「レポート」「手紙」といった、ツリー構造の最上位の階層のエレメントを一覧表示する。

【0022】ユーザは、画面に一覧表示されたDTDの中から、検索する文書の型に合ったDTDとして、例えば「レポート」を選択して、スタイル情報を表示すべきDTDを決定する(ステップS1)。DSSSLエンジン8は、

決定されたDTD、本例では「レポート」の文書成形の規則を定めたスタイルシートをスタイルシートDB9から読み込み(ステップS2)、フローオブジェクトを生成して、図2に示すような「レポート」の文書フォーマット図を表示装置1に表示する(ステップS3)。

【0023】この文書フォーマット図により、ユーザはその文書構造を知らなくても、「レポート」が、「宛先」「発信元」「タイトル」「本文」及び「担当」の要素から構成され、「本文」が「概要」及び「詳細」の要素から構成されていることと、それぞれの要素の内容の表示されるレイアウト位置がどの辺りであるかを、その枠組みで視覚的に把握することができる。

【0024】インタフェースモジュール3は、表示された文書フォーマット図に対してキーワードの入力があるか否かを監視し(ステップS4)、キーワードが入力された場合は、その入力位置から、キーワードが入力された要素を判定する。インタフェースモジュール3は、入力されたキーワードからクエリを作成してサーチエンジン4に渡す(ステップS5)。サーチエンジン4はインタフェースモジュール3から渡されたクエリにより、ドキュメントDB7の中の検索対象の要素の内容を検索して目的の文書を獲得する(ステップS6)。

【0025】このとき、図4に示すように、複数の要素に対するキーワードの入力が可能である。複数の要素に対してキーワードが入力された場合、インタフェースモジュール3は、入力された複数のキーワードとデフォルトで設定されている検索条件(OR検索又はAND検索)とからクエリを作成するか、又は、図4に示すように、検索条件の入力画面を表示してユーザに選択を促し、選択された検索条件と入力されたキーワードとからクエリを作成する。

【0026】例えば、「タイトル」要素にキーワード「見積もり」が、また「担当」要素にキーワード「富士太郎」が入力され、AND検索を行う場合は、「タイトル」要素の中に文字列「見積もり」が含まれるとともに、「担当」要素の中に文字列「富士太郎」が含まれる文書を検索する。上述の入力でOR検索を行う場合は、要素「タイトル」の中に文字列「見積もり」が含まれるか、又は要素「担当」の中に文字列「富士太郎」が含まれる文書を検索する。

【0027】図4は、検索条件の指定において要素間の関係(and, or)を表したリンクを別のリンクの対象とする場合の入力例を示している。「宛先」要素にキーワード「経理」、「発信元」要素にキーワード「経理」、「タイトル」要素にキーワード「見積もり」、「担当」要素にキーワード「富士太郎」が入力されている。

【0028】入力したキーワードによって、「AND検索」「OR検索」「複合検索」のいずれを実行するかをユ

ーザに選択させるために、例えばボタン、メニュー等を表示する。さらに「複合検索」においては「ANDリンク」「ORリンク」のいずれを設定するかをユーザに選択させるためのボタン、メニュー等を表示する。

【0029】例えば、ユーザが、ボタン、メニュー等の「複合検索」及び「ANDリンク」を選択し、ポインタPによって「発信元」、「タイトル」、及び「担当」要素を指定すると、これらの要素が、例えば実線で結ばれ、ANDリンクが設定されたことが表示される。次に、ユーザが「ORリンク」を選択し、ANDリンクの実線と、「宛先」要素をポインタPによって指定すると、ANDリンクの実線と「宛先」要素とが、例えば破線で結ばれ、ORリンクが設定されたことが表示される。

【0030】以上の設定の結果、「発信元」要素に文字列「経理」が含まれ、かつ「タイトル」要素に文字列「見積もり」が含まれ、かつ「担当」要素に文字列「富士太郎」が含まれる文書か、又は「宛先」要素に文字列「経理」が含まれる文書を検索するクエリが作成される。

【0031】さらに、検索条件として、キーワードが入力された要素だけを検索対象とする「単体検索」か、その要素の下位の階層の要素も検索対象に含む「階層検索」かを選択することもできるようになっている。

【0032】なお、検索条件の選択方法はメニュー、ボタンに限定されるものではなく、文字入力、番号選択等の他の選択方法も可能である。また、複合検索におけるANDリンク、ORリンクの別を表す方法は実線、破線等の線種による方法に限らず、線の色を変える、AND/ORの文字で表示する等の他の方法であってもよい。さらに、必ずしも表示しなくてもよい。

【0033】次に、本発明方法に要素構造情報のみを用いる場合について、図5のキーワード入力の説明図、図6のフローチャート、及び図7のキーワード入力例の図に基づいて説明する。前述の場合と同様に、インタフェースモジュール3は、ドキュメントDB7に格納されているSGML文書のDTDを読み出し、「レポート」「手紙」といった、ツリー構造の最上位の階層の一覧を表示するが、本例ではリスト形式で表示する(ステップS11)。

【0034】ユーザは、画面にリスト形式で一覧表示されたDTDの中から、検索する文書の型に合ったDTDを決定する(ステップS12)。このとき、インタフェースモジュール3は、例えば「レポート」の文字列上がクリックされたか、又は「レポート」の欄に設けられた、下位の階層が存在することを示すマークがクリックされたかを識別して、要素が選択されたか否かを判定する(ステップS13)。要素の欄のマークがクリックされた場合、「レポート」の下位の階層の「宛先」「タ



イトル」…のエレメント名(タグ名)をリスト形式で表示する(ステップS14)。

【0035】「レポート」の階層の場合と同様、下位の階層のエレメントの文字列上がクリックされたか、又はエレメントの欄に設けられているマークがクリックされたかを識別して、エレメントが選択されたか否かを判定する(ステップS13)。例えば、「本文」のエレメントの欄のマークがクリックされた場合は、さらに下位の階層のタグリストを表示する(ステップS14)。

【0036】一方、エレメントの文字列がクリックされた場合は、キーワードの入力画面を表示し、カーソルを点滅させてキーワードの入力待ちとなる(ステップS15)。キーワードが入力されると、検索の開始が指示されたか否かをチェックし(ステップS16)、検索の開始が指示されない場合はステップS13に戻って、ユーザからの次の入力待ちになる。

【0037】前述の場合と同様、入力したキーワードによって、「AND 検索」「OR検索」「複合検索」のいずれを実行するかをユーザに選択させるために、例えばボタン、メニュー等を表示する。さらに「複合検索」においては「AND リンク」「ORリンク」のいずれを設定するかをユーザに選択させるためのボタン、メニュー等を表示する。また、前述の場合と同様、検索条件として、キーワードが入力されたエレメントだけを検索対象とする

「単体検索」か、そのエレメントの下位の階層のエレメントも検索対象に含む「階層検索」かを選択することもできるようになっている。

【0038】検索の開始が指示されると、インタフェースモジュール3は、入力されたキーワードからクエリを作成して(ステップS17)サーチエンジン4に渡し、サーチエンジン4はインタフェースモジュール3から渡されたクエリにより、ドキュメントDB7の中の検索対象のエレメントの内容を検索して目的の文書を獲得する(ステップS18)。

【0039】このとき、図7に示すように、複数のエレメントに対するキーワードの入力が可能である。複数のエレメントに対してキーワードが入力された場合、インタフェースモジュール3は、入力された複数のキーワードとデフォルトで設定されている検索条件(OR検索又はAND 検索)とからクエリを作成するか、又は、図7に示すように、検索条件の入力画面を表示してユーザに選択を促し、選択された検索条件と入力されたキーワードとからクエリを作成する。

【0040】図7のようなキーワード入力で、AND 検索と単体検索とが選択された場合、「レポート」エレメントに文字列「経理」が含まれ、かつ「タイトル」エレメントに文字列「見積もり」が含まれ、かつ「本文」エレメントに文字列「FMV」が含まれる文書を検索するクエリが作成される。

【0041】また図7のようなキーワード入力で、OR検

索と単体検索とが選択された場合、「レポート」エレメントに文字列「経理」が含まれるか、又は「タイトル」エレメントに文字列「見積もり」が含まれるか、又は「本文」エレメントに文字列「FMV」が含まれる文書を検索するクエリが作成される。なお、各エレメントに与えるキーワードにAND、OR、正規表現を記述することもできる。

【0042】図7のようなキーワード入力で、AND 検索と階層検索とが選択された場合、「レポート」エレメントより下位のエレメントに文字列「経理」が含まれ、かつ「タイトル」エレメントに文字列「見積もり」が含まれ、かつ「本文」エレメントより下位のエレメントに文字列「FMV」が含まれる文書を検索するクエリが作成される。

【0043】以上のように、本例の場合は階層順を追ってエレメントの構造図を表示していくので、一画面でエレメントの構造を表示できない巨大な構造の構造化文書に対しても、目的のエレメントにキーワードをスムーズに与えることができる。

【0044】次に、本発明方法にエレメント構造情報のみを用いる他の場合について、図8のキーワード入力の説明図、図9のフローチャート、及び図10のキーワード入力例の図に基づいて説明する。前述の場合と同様に、インタフェースモジュール3は、ドキュメントDB7に格納されているSGML文書のDTDを読み出し「レポート」「手紙」といった、ツリー構造の最上位の階層のエレメントを一覧表示する。

【0045】ユーザは、画面にリスト形式で一覧表示されたDTDの中から、検索する文書の型に合ったDTDを決定する(ステップS21)。DSSSL エンジン8は、決定されたDTD、本例では「レポート」の文書成形の規則を定めたスタイルシートをスタイルシートDB9から読み込み、フローオブジェクトを生成して、図8に示すような、「レポート」のツリー構造の構造図を表示装置1に表示する(ステップS22)。

【0046】この構造図により、ユーザはその文書構造を知らなくても、「レポート」が、「宛先」「発信元」「タイトル」「本文」及び「担当」のエレメントから構成され、「本文」が「概要」及び「詳細」のエレメントから構成されていることを把握することができる。

【0047】インタフェースモジュール3は検索の開始が指示されたか否かをチェックし(ステップS23)、検索開始が指示されない場合は、表示した構造図のエレメントに対する選択入力待ちとなる(ステップS24)。いずれかのエレメントが選択されると、そのエレメントに対応付けて、キーワードの入力画面を表示し、カーソルを点滅させてキーワードの入力待ちとなる(ステップS25)。

【0048】キーワードが入力されると、検索の開始が指示されたか否かをチェックし(ステップS23)、検索

の開始が指示されない場合は、表示した構造図のエレメントに対する選択入力待ちとなる（ステップ S 24）。いずれかのエレメントが選択されると、そのエレメントに対応付けて、キーワードの入力画面を表示し、カーソルを点滅させてキーワードの入力待ちとなる（ステップ S 25）。

【0049】検索の開始が指示されると、インタフェースモジュール 3 は、入力されたキーワードからクエリを作成してサーチエンジン 4 に渡し（ステップ S 26）、サーチエンジン 4 はインタフェースモジュール 3 から渡されたクエリにより、ドキュメント DB 7 中の検索対象のエレメントの内容を検索して目的の文書を獲得する（ステップ S 27）。

【0050】前述の場合と同様、入力したキーワードによって、「AND 検索」「OR 検索」「複合検索」のいずれを実行するかをユーザに選択させるために、例えばボタン、メニュー等を表示する。さらに「複合検索」においては「AND リンク」「OR リンク」のいずれを設定するかをユーザに選択させるためのボタン、メニュー等を表示する。

【0051】また、前述の場合と同様、検索条件として、キーワードが入力されたエレメントだけを検索対象とする「単体検索」か、そのエレメントの下位の階層のエレメントも検索対象に含む「階層検索」かを選択することもできるようになっている。なお、各エレメントに与えるキーワードに AND、OR、正規表現を記述することができる。

【0052】このとき、図 10 に示すように、複数のエレメントに対するキーワードの入力が可能である。複数のエレメントに対してキーワードが入力された場合、インタフェースモジュール 3 は、入力された複数のキーワードとデフォルトで設定されている検索条件（OR 検索又は AND 検索）とからクエリを作成するか、又は、図 10 に示すように、検索条件の入力画面を表示してユーザに選択を促し、選択された検索条件と入力されたキーワードとからクエリを作成する。

【0053】図 10 のようなキーワード入力で、AND 検索と単体検索とが選択された場合、「タイトル」エレメントに文字列「見積み」が含まれ、かつ「担当」エレメントに文字列「富士太郎」が含まれる文書を検索するクエリが作成される。また図 10 のようなキーワード入力 40 で、OR 検索と単体検索とが選択された場合、「タイトル」エレメントに文字列「見積み」が含まれるか、又は「担当」エレメントに文字列「富士太郎」が含まれる

文書を検索するクエリが作成される。

【0054】なお、上述のようなキーワード入力のいずれかをユーザに選択させるメニュー画面を設けてもよい。

【0055】

【発明の効果】以上のように、本発明の構造化文書データベースの検索方法、検索システム及び記録媒体は、構造化文書の文書構造の図式をエレメント名を用いてユーザに示し、図式上のエレメント名に対してユーザがキーワードを入力したエレメントを対象として検索するので、文書構造がユーザにとって未知である構造化文書を検索する場合であっても、検索対象のエレメントの指定、及びそのエレメントに対するキーワードの入力が容易であって、検索が短時間でできるという優れた効果を奏する。

【図面の簡単な説明】

【図 1】本発明のシステム構成図である。

【図 2】スタイル情報を用いたキーワード入力の説明図である。

20 【図 3】図 2 による検索方法のフローチャートである。

【図 4】スタイル情報を用いたキーワード入力例の図である。

【図 5】構造情報を用いたキーワード入力の説明図（その 1）である。

【図 6】図 5 による検索方法のフローチャートである。

【図 7】図 5 の構造情報を用いたキーワード入力例の図である。

【図 8】構造情報を用いたキーワード入力の説明図（その 2）である。

【図 9】図 8 による検索方法のフローチャートである。

【図 10】図 8 の構造情報を用いたキーワード入力例の図である。

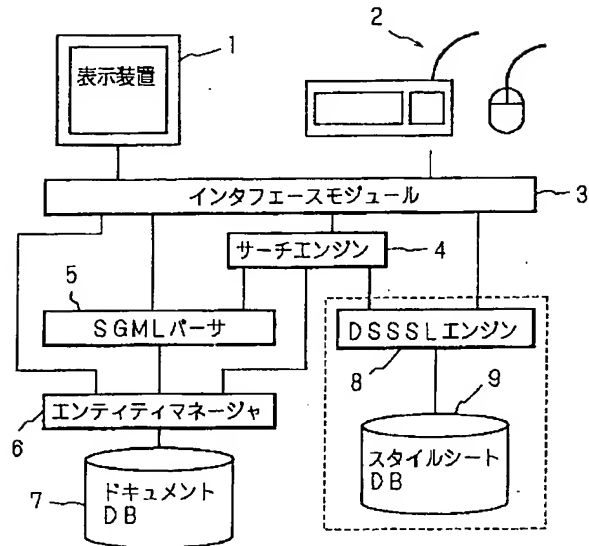
【図 11】構造化文書の構造の一例の図である。

【符号の説明】

- 1 表示装置
- 2 入力装置
- 3 インタフェースモジュール
- 4 サーチエンジン
- 5 SGML パーサ
- 6 エンティティマネージャ
- 7 ドキュメント DB
- 8 DSSSL エンジン
- 9 スタイルシート DB

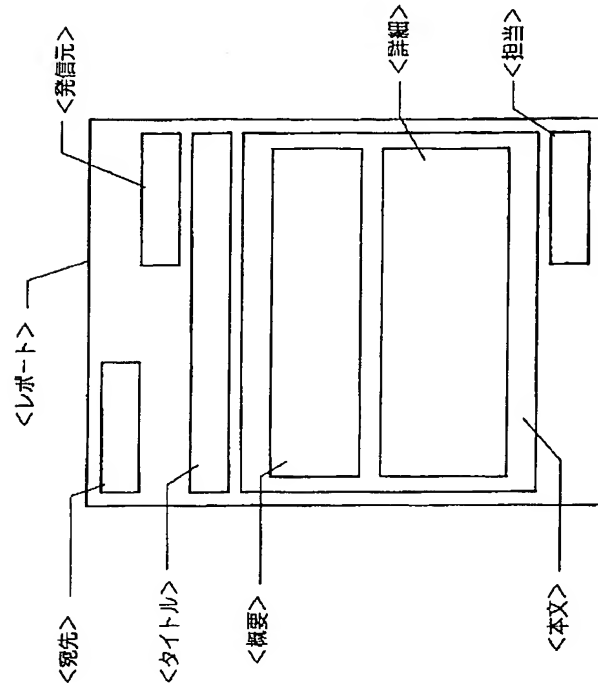
【図1】

本発明のシステム構成図



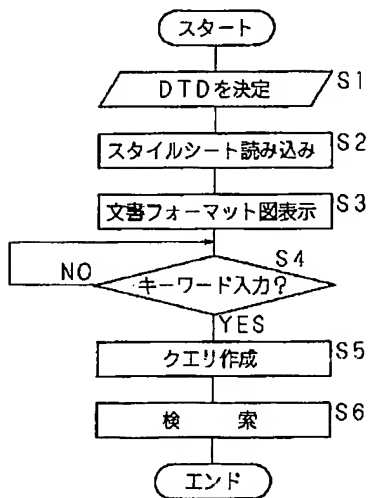
【図2】

スタイル情報を用いたキーワード入力の説明図



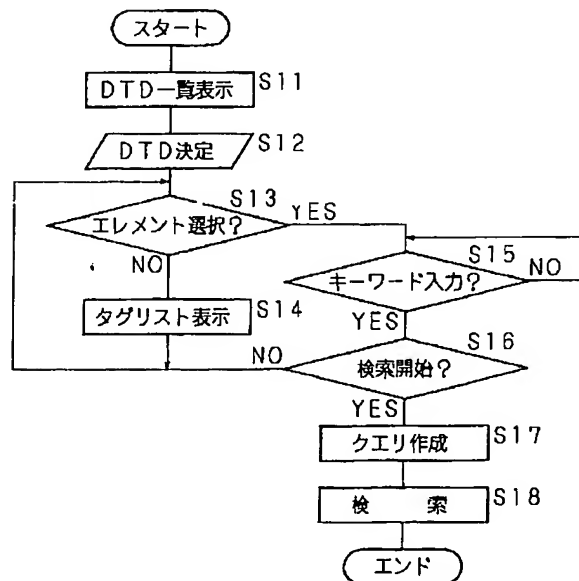
【図3】

図2による検索方法のフローチャート



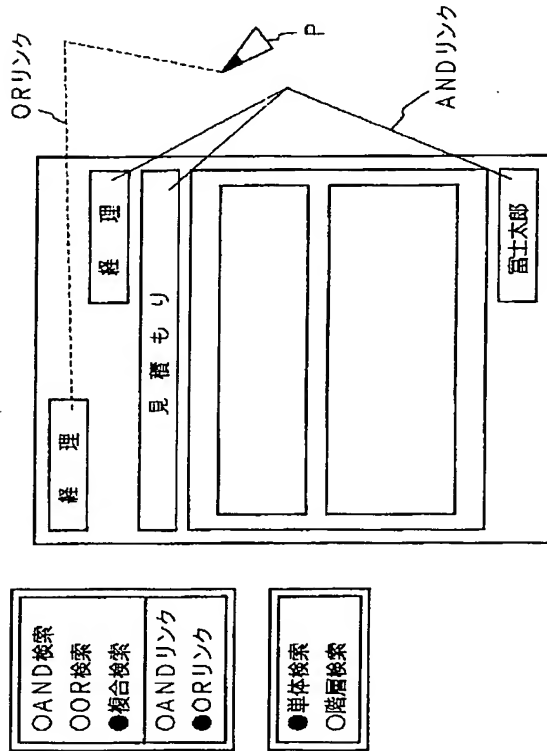
【図6】

図5による検索方法のフローチャート



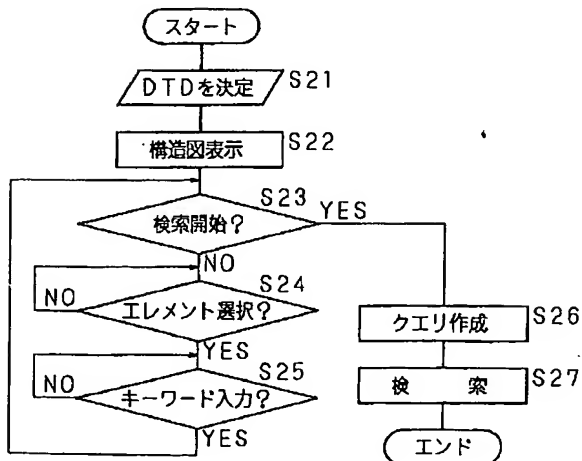
【図 4】

スタイル情報を用いたキーワード入力例の図



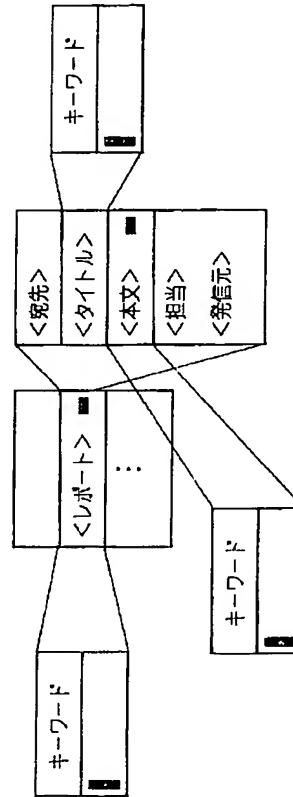
【図 9】

図 8 による検索方法のフローチャート



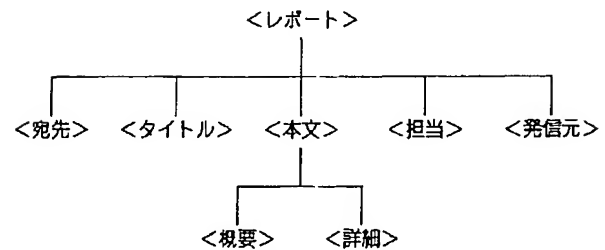
【図 5】

構造化情報を用いたキーワード入力の説明図(その1)



【図 11】

構造化文書の構造の一例の図





【図 1 0】

図8の構造情報を用いたキーワード入力例の図

